Vector Autoregressive models

Rishman Jot Kaur Chahal

Mehta Family School of Data Science and Artificial Intelligence
Indian Institute of Technology Roorkee

October 5, 2023

Vector Autoregressive Models

Vector → $y_{1t}, y_{2t}, \dots$

t	y_t
⋮	⋮
	AR

- It is an evolved technique and has been applied in various fields of economics and finance.

$$y_t = \{y_{1t}, y_{2t}\}.$$

- There are different versions of VAR
  - Unrestricted VAR
  - Restricted/Structural VAR
  - Sign-restriction VAR
  - Bayesian VAR
  - Markov-Switching VAR (MSVAR)
  - There is also Panel VAR

$$y_{1t} = f(y_{1t-1}, y_{2t})$$

$$y_{2t} = f(y_{1t}, y_{2t-1})$$

→ Endogenous.

- A natural generalisation of univariate autoregressive models popularised by Sims (1980) as discussed earlier.
- A VAR is in a sense a systems regression model i.e. there is more than one dependent variable.

$y_{1t}, y_{2t}, \ldots$

- It is a hybrid between univariate time series models and simultaneous equations model. VARs have often been advocated as an alternative to large-scale simultaneous equations structural models.
- What is a simultaneous model and how does it looks like???

- All the structural equations that we have considered till now have been single equations model:

$$y = X\beta + u$$

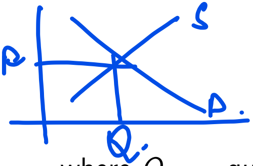
X = Exog.
Y = Endog

- Remember that in such models X is assumed as exogenous and causality runs from X to y and not vice-versa i.e., that changes in the values of the explanatory variables cause changes in the values of y, but that changes in the value of y will not impact upon the explanatory variables.

- Also, y is an endogenous variable which is determined by the equation above.

- What will happen if this situation violates? How will that happen?

- Consider the following 2 equations that describe a possible model for the total aggregate (country-wide) supply of new houses (or any other physical asset).



$$Q_{dt} = \alpha + \beta P_t + \gamma S_t + u_t \tag{1}$$

$$Q_{st} = \lambda + \mu P_t + \tau T_t + v_t \tag{2}$$

$$Q_{dt} = Q_{st} \tag{3}$$

P =

where Q_{dt} = quantity of new houses demanded at time t, Q_{st} = quantity of new houses supplied (built) at time t, P_t = (average) price of new houses prevailing at time t, S_t = price of substitute (e.g. older houses) and T_t = some variable embodying the state of housebuilding technology, u_t and v_t are error terms.

- Eq. 1 is the equation for modelling the demand for new houses, Eq. 2 models the supply and Eq. 3 is an equilibrium condition.

$Q_D = Q_S$.

- Assuming that the market always clears, that is, that the market is always in equilibrium, and dropping the time subscripts for simplicity, we can write the equations as:

$$Q = \alpha + \beta P + \gamma S + u \tag{4}$$

$$Q = \lambda + \mu P + \tau T + v \tag{5}$$

$Q = \alpha + \beta P + \gamma S + u = \lambda + \mu + \tau T + v$.

- Eq. 4 and 5 together comprise a simultaneous structural form of the model, or a set of structural equations. $P = \_\_\_\_\_\_$

- The point is that price and quantity are determined simultaneously (price affects quantity and quantity affects price).

- But, We cannot estimate equation 4 and 5 validly using OLS. Why?

- Because both 4 and 5 are related to each other as they both contain P and Q and OLS require them to be estimated separately.
- So, we can estimate them separately using OLS. No! Both depends upon P. How?
- Lets just look at their reduced forms.

- Solving for Q:

$$\alpha + \beta P + \gamma S + u = \lambda + \mu P + \tau T + v \tag{6}$$

- Solving for P:

$$\frac{Q}{\beta} - \frac{\alpha}{\beta} - \frac{\gamma S}{\beta} - \frac{u}{\beta} = \frac{Q}{\mu} = \frac{\lambda}{\mu} - \frac{\tau T}{\mu} - \frac{v}{\mu} \tag{7}$$

- Rearrange eq. 6

$P = f(u,v).$

random

endog.

$$\beta P - \mu P = \lambda - \alpha + \tau T - \gamma S + v - u \tag{8}$$

$$(\beta - \mu)P = (\lambda - \alpha) + \tau T - \gamma S + (v - u) \tag{9}$$

$$P = \frac{\lambda - \alpha}{\beta - \mu} + \frac{\tau}{\beta - \mu}T - \frac{\gamma}{\beta - \mu}S + \frac{v - u}{\beta - \mu} \tag{10}$$

- Now, remember one of the CLRM assumptions was that X and u are independent and also $E(u) = 0$ and $E(X'u) = 0$. But eq. 10 shows that P is related to the errors in eq. 4 and 5 and thus P is stochastic.

- Thus, the assumption is violated.

- What will happen if we ignore this?

Reg. $\hat{\beta}$, $Y = X\beta + u$

$E(\hat{\beta}) = \beta$.

- Remember our $\hat{\beta} = (X'X)^{-1}X'y$

- Replacing y in the RHS will give us

$$\hat{\beta} = (X'X)^{-1}X'(X\beta + u)$$

- Thus, $\hat{\beta} = \beta + (X'X)^{-1}X'u$. Taking expectations on both sides.

$$E(\hat{\beta}) = \beta + \underbrace{E((X'X)^{-1}X'u)}_{\text{bias}}$$

bias.

- If the Xs are non-stochastic (i.e., if the assumption had not been violated), $E[(X'X)^{-1}X'u] = (X'X)^{-1}X'E[u] = 0$ and thus $E(\hat{\beta}) = \beta$ indicating $\hat{\beta}$ would be unbiased.
- This will not be a situation now as equation is part of a system and $E[(X'X)^{-1}X'u] \neq 0$. simultaneity bias.
- Thus, OLS application to structural equations will lead to biased coefficient estimates. This is known as simultaneity bias or simultaneous equations bias.
- There exists various estimation procedures that can be used to estimate such models like Indirect Least squares ILS and two-stage least squares (2SLS) and Instrument Variables (IV).

- Lets come back to our discussion of VAR models.
- Simplest case is a bivariate VAR Y_{1t}, Y_{2t}.

$$y_{1t} = \beta_{10} + \beta_{11}y_{1t-1} + \dots + \beta_{1k}y_{1t-k} + \alpha_{11}y_{2t-1} + \dots + \alpha_{1k}y_{2t-k} + u_{1t}$$
$$y_{2t} = \beta_{20} + \beta_{21}y_{2t-1} + \dots + \beta_{2k}y_{2t-k} + \alpha_{21}y_{1t-1} + \dots + \alpha_{2k}y_{1t-k} + u_{2t}$$

where u_{it} is an iid disturbance term with
$E(u_{it}) = 0, i = 1, 2; E(u_{1t}u_{2t}) = 0$. No Autocorr.

- The analysis could be extended to a VAR(g) model, or so that there are g variables $y_{1t}, y_{2t}, \dots, y_{gt}$ and g equations. $Y_{1t} = f(Y_{1t}, Y_{2t})$
$Y_{2t} = f(Y_{1t}, Y_{2t})$
- As should already be evident, an important feature of the VAR model is its flexibility and the ease of generalisation. For example, the model could be extended to encompass moving average errors, which would be a multivariate version of an ARMA model, known as a VARMA.

- One important feature of VARs is the compactness with which we can write the notation. For example, consider the case from above where k=1. (Lags)

Restrictions.

- We can write this as

$y_{1t} = \beta_{10} + \beta_{11} y_{1t-1} + \alpha_{11} y_{2t-1} + u_{1t}$

$y_{2t} = \beta_{20} + \beta_{21} y_{2t-1} + \alpha_{21} y_{1t-1} + u_{2t}$ $+ \alpha_{22} y_{1t-2}$

or

$$\begin{pmatrix} y_{1t} \\ y_{2t} \end{pmatrix} = \begin{pmatrix} \beta_{10} \\ \beta_{20} \end{pmatrix} + \begin{pmatrix} \beta_{11} & \alpha_{11} \\ \alpha_{21} & \beta_{21} \end{pmatrix} \begin{pmatrix} y_{1t-1} \\ y_{2t-1} \end{pmatrix} + \begin{pmatrix} u_{1t} \\ u_{2t} \end{pmatrix}$$

or even more compactly as

$$\left(\begin{array}{c} y_t \\ gx1 \end{array} \right) = \left(\begin{array}{c} \beta_0 \\ gx1 \end{array} \right) + \left(\begin{array}{c} \beta_1 \\ gxg \end{array} \right) \left(\begin{array}{c} y_{t-1} \\ gx1 \end{array} \right) + \left(\begin{array}{c} u_t \\ gx1 \end{array} \right)$$

Vector Autoregressive Models Compared with Structural Equations Models

- Advantages of VAR Modelling
  - Do not need to specify which variables are endogenous or exogenous - all are endogenous
  - Allows the value of a variable to depend on more than just its own lags or combinations of white noise terms, so more general than ARMA modelling
  - Forecasts are often better than "traditional structural" models.
- Problems with VAR's
  - VAR's are a-theoretical (as are ARMA models)
  - How do you decide the appropriate lag length?

  vector AIC, BIC.
  - So many parameters! If we have g equations for g variables and we have k lags of each of the variables in each equation, we have to estimate $(g + kg^2)$ parameters. e.g. g=3, k=3, parameters = 30.
  - Do we need to ensure all components of the VAR are stationary?
  - How do we interpret the coefficients?

Choosing the Optimal Lag Length for a VAR

- 2 possible approaches: cross-equation restrictions and information criteria. AIC, BIC.

Cross-Equation Restrictions Restricted VAR.

- In the spirit of (unrestricted) VAR modelling, each equation should have the same lag length. A VAR with different lag lengths for each equation could be viewed as a restricted VAR. How???

- Consider a VAR with three lags of both variables in one equation and four lags of each variable in the other equation.
    - This could be viewed as a restricted model where the coefficient on the fourth lags of each variable in the first equation have been set to zero.

Information Criteria for VAR Lag Length Selection

- Multivariate versions of the information criteria are required. These can be defined as:

$$MAIC = ln|\hat{\sum}| + 2k^{'}/T$$
$$MSBIC = ln|\hat{\sum}| + \frac{k^{'}}{T}ln(T)$$
$$MHQIC = ln|\hat{\sum}| + \frac{2k^{'}}{T}ln(ln(T))$$

where all notation is as above and $k^{'}$ is the total number of regressors in all equations, which will be equal to $g^2k + g$ for g equations, each with k lags of the g variables, plus a constant term in each equation. The values of the information criteria are constructed for 0, 1, ... lags (up to some prespecified maximum $\bar{k}$).

- It is likely that, when a VAR includes many lags of variables, it will be difficult to see which sets of variables have significant effects on each dependent variable and which do not. For illustration, consider the following bivariate VAR(3):

$$\begin{pmatrix} y_{1t} \\ y_{2t} \end{pmatrix} = \begin{pmatrix} \alpha_{10} \\ \alpha_{20} \end{pmatrix} + \begin{pmatrix} \beta_{11} & \beta_{12} \\ \beta_{21} & \beta_{22} \end{pmatrix} \begin{pmatrix} y_{1t-1} \\ y_{2t-1} \end{pmatrix}$$

$$+ \begin{pmatrix} \gamma_{11} & \gamma_{12} \\ \gamma_{21} & \gamma_{22} \end{pmatrix} \begin{pmatrix} y_{1t-2} \\ y_{2t-2} \end{pmatrix} + \begin{pmatrix} \delta_{11} & \delta_{12} \\ \delta_{21} & \delta_{22} \end{pmatrix} \begin{pmatrix} y_{1t-3} \\ y_{2t-3} \end{pmatrix} + \begin{pmatrix} u_{1t} \\ u_{2t} \end{pmatrix}$$

Block Significance and Causality Tests (Cont'd)

$$y_{1t} = \alpha_{10} + \beta_{11} y_{1,t-1} + \beta_{12} y_{2,t-1} + \gamma_{11} y_{1,t-2} + \gamma_{12} y_{2,t-2} + \delta_{11} y_{1,t-3} + \delta_{12} y_{2,t-3} + u_{1,t}$$

$$y_{2t} = \alpha_{20} + \beta_{21} y_{1,t-1} + \beta_{22} y_{2,t-1} + \gamma_{21} y_{1,t-2} + \gamma_{22} y_{2,t-2} + \delta_{21} y_{1,t-3} + \delta_{22} y_{2,t-3} + u_{2,t}$$

- We might be interested in testing the following hypotheses, and their implied restrictions on the parameter matrices:

Hypothesis	Implied Restriction
1. Lags of y_{1t} do not explain current y_{2t}	$\beta_{21} = 0$ and $\gamma_{21} = 0$ and $\delta_{21} = 0$
2. Lags of y_{1t} do not explain current y_{1t}	$\beta_{11} = 0$ and $\gamma_{11} = 0$ and $\delta_{11} = 0$
3. Lags of y_{2t} do not explain current y_{1t}	$\beta_{12} = 0$ and $\gamma_{12} = 0$ and $\delta_{12} = 0$
4. Lags of y_{2t} do not explain current y_{2t}	$\beta_{22} = 0$ and $\gamma_{22} = 0$ and $\delta_{22} = 0$

- Each of these four joint hypotheses can be tested within the F-test framework, since each set of restrictions contains only parameters drawn from one equation.
- These tests could also be referred to as Granger causality tests.
- Granger causality tests seek to answer questions such as "Do changes in y_1 cause changes in y_2 ?" If y_1 causes y_2 , lags of y_1 should be significant in the equation for y_2 . If this is the case, we say that y_1 "Granger-causes" y_2.
- If y_2 causes y_1 , lags of y_2 should be significant in the equation for y_1..
- If both sets of lags are significant, there is "bi-directional causality"

$Y_1 \to Y_2$ & $Y_2 \to Y_1$. $\Rightarrow$ Bi-directional causality.

Impulse Responses

- VAR models are often difficult to interpret: one solution is to construct the impulse responses and variance decomposition.
- Impulse responses trace out the responsiveness of the dependent variables in the VAR to shocks to the error term. A unit shock is applied to each variable and its effects are noted.
- Consider for example a simple bivariate VAR(1):

  $$y_{1t} = \beta_{10} + \beta_{11} y_{1t-1} + \alpha_{11} y_{2t-1} + \alpha_{12} y_{2t} + u_{1t}$$
  $$y_{2t} = \beta_{20} + \beta_{21} y_{2t-1} + \alpha_{21} y_{1t-1} + \alpha_{22} y_{1t} + u_{2t}$$

  unit shock.

- A change in u_{1t} will immediately change y_1 . It will change change y_2 and also y_1 during the next period.
- We can examine how long and to what degree a shock to a given equation has on all of the variables in the system.

- Consider a simple bivariate VAR(1) model:

$$y_t = A_1 y_{t-1} + u_t$$

where A_1 is

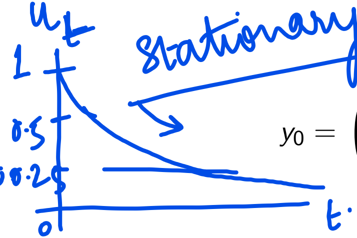
$$\begin{pmatrix} 0.5 & 0.3 \\ 0.0 & 0.2 \end{pmatrix}$$

- VAR can further be written as

$$\begin{pmatrix} y_{1t} \\ y_{2t} \end{pmatrix} = \begin{pmatrix} 0.5 & 0.3 \\ 0.0 & 0.2 \end{pmatrix} \begin{pmatrix} y_{1t-1} \\ y_{2t-1} \end{pmatrix} + \begin{pmatrix} u_{1t} \\ u_{2t} \end{pmatrix}$$

- Consider the effect at time t = 0, 1, . . . , of a unit shock to y1t at time t = 0

Impulse Response: Example



Stationary.

$$y_0 = \begin{pmatrix} y_{10} \\ y_{20} \end{pmatrix} = \begin{pmatrix} u_{10} \\ u_{20} \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix}$$

$$y_1 = A_1 y_0 = \begin{pmatrix} 0.5 & 0.3 \\ 0.0 & 0.2 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 0.5 \\ 0 \end{pmatrix}$$

$$y_2 = A_1 y_1 = \begin{pmatrix} 0.5 & 0.3 \\ 0.0 & 0.2 \end{pmatrix} \begin{pmatrix} 0.5 \\ 0 \end{pmatrix} = \begin{pmatrix} 0.25 \\ 0 \end{pmatrix}$$

And so on. It would thus be possible to plot the impulse response functions of y_{1t} and y_{2t} to a unit shock in y_{1t}.

- It tells how much of a change in a variable is due to its own shock and how much due to shocks to other variables. In the SR most of the variation is due to own shock. But as the lagged variables effect starts kicking in, the percentage of the effect of other shocks increases over time.
- We want to calculate the n-period forecast error of x in order to find that of say, y.

Variance Decomposition

- Impulse responses + Variance decomposition = innovation accounting.

- Variance decomposition's offer a slightly different method of examining VAR dynamics. They give the proportion of the movements in the dependent variables that are due to their "own" shocks, versus shocks to the other variables.
- This is done by determining how much of the s-step ahead forecast error variance for each variable is explained innovations to each explanatory variable ($s = 1,2,\ldots$).
- The variance decomposition gives information about the relative importance of each shock to the variables in the VAR.

Impulse Responses and Variance Decompositions: The Ordering of the Variables

- But for calculating impulse responses and variance decompositions, the ordering of the variables is important.
- The main reason for this is that above, we assumed that the VAR error terms were statistically independent of one another.
- This is generally not true, however. The error terms will typically be correlated to some degree.
- Therefore, the notion of examining the effect of the innovations separately has little meaning, since they have a common component.
- What is done is to "orthogonalise" the innovations.
- In the bivariate VAR, this problem would be approached by attributing all of the effect of the common component to the first of the two variables in the VAR.
- In the general case where there are more variables, the situation is more complex but the interpretation is the same.

Thank you!!